DOCUMENT RESUME

ED 373 552 FL 022 352

AUTHOR Cutting, Joan
TITLE A Coding System for Analysing a Spoken Text Database.
PUB DATE 94
NOTE 16p.; For serial publication in which this paper appears, see FL 022 351.
PUB TYPE Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT Edinburgh Working Papers in Applied Linguistics; n5 p1-14 1994

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Applied Linguistics; *Coding; Databases; *Discourse Analysis; Foreign Countries; *Graduate Students; Grammar; *Group Discussion; Group Dynamics; Higher Education; *Language Patterns; Language Research; *Language Styles; Peer Relationship; Speech Acts
IDENTIFIERS University of Edinburgh (Scotland)

ABSTRACT

This paper describes a coding system devised to analyze conversations of graduate students in applied linguistics at Edinburgh University. The system was devised to test the hypothesis that as shared knowledge among conversation participants grows, the textual density of in-group members has more cues than that of strangers. The informal conversations of 6 students who shared classes in linguistics were recorded over the course of the 1991-92 academic year. The paper provides details of the lexical and grammatical tags applied to the texts of the conversations and outlines the code for the analysis of each discourse unit's assumed knowledge area, macro-function, speech act, and move. The problems encountered in the implementation of this system are discussed and explained. (Contains 15 references.) (MDM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 552

A Coding System for Analysing a Spoken Text Database

Joan Cutting (DAL)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Brian Parkinson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

FL 022 352

A CODING SYSTEM FOR ANALYSING A SPOKEN TEXT DATABASE

Joan Cutting (DAL)

Abstract

This paper describes a coding system devised to analyse conversations of 1991-92 Applied Linguistics MSc students at Edinburgh University. It gives details of the lexical and grammatical tags that are applied to the text itself and outlines the code for the analysis of each discourse unit's assumed knowledge area, macro-function, speech act, and move. The problems encountered in the implementation of this system are discussed and explained, and solutions are offered.

1. Introduction

There have been several studies, in the fields of sociolinguistics and psycholinguistics, that compare the language of strangers and that of friends (Tannen 1989; Duck 1991). However, it would appear that there is no longitudinal linguistic study that follows through the interactions of speakers from the moment that they first meet until they become friends or associates, in order to discover exactly how their language develops over time.

I am making a developmental study of casual conversations of MSc students from the 1991-92 Edinburgh University Applied Linguistics course, aiming to find exactly how their language changes as they form a discourse community. The analysis is hoped to explain how the in-group's code becomes increasingly restricted (Bernstein 1971) and thus inaccessible to an outsider to this MSc group, and to provide a model for predicting language changes during discourse community formation. It is not an *a priori* model: the categories have been devised as a result of examining the dialogues with an ethnomethodological eye.

In order that the pragmatic analysis of the implicit language that evolves in this closed network academic group (Levy 1979; Kreckel 1981) can be both qualitative and quantitative, I have devised a detailed coding system. The system contains certain lexical and grammatical features that depend on the context of the MSc course for their full meaning: special terms and names, general words (Halliday and Hasan 1976) and exophoric reference, substitution and ellipsis. I call these features implicit 'contextualisation cues', to use Gumperz's term for linguistic features that contribute to the 'signalling of contextual presuppositions' (1982:71). I hypothesise that over time, as the group becomes closer knit, united by the common experience of the MSc course, there will be an increase in implicit contextualisation cues.

The coding system also contains functional categories, since the linguistic features must be seen in the context of when, how and why they are used, so that a statement may be made about group dynamics. Thus I analyse each discourse unit in terms of perlocutionary function, implicature and speech act, as well as the knowledge area assumed.

This paper describes each category of my coding system. It discusses the difficulties involved in making the system work and adapting the categories to meet the needs of natural language.

2. The Overall Methodology

2.1 Hypotheses

I hypothesise that as shared knowledge grows, the textual density of implicit contextualisation cues increases and that the language of in-group members has more cues than that of strangers. I hypothesise that course-related topics will be more impenetrable than non-course-related topics to an outsider to the course; that course-related topics will become more frequent than non-course-related topics; and that this will cause the conversations to have larger impenetrable sections because of both the assumed knowledge area and the density of implicit contextualisation cues.

My hypothesis about the par[illegible] development of the implicit contextualisation cues is that after the beginning [illegible] the MSc course, there will be a peak of special terms (eg. 'X-bar'), proper names (eg. 'Chomsky'), demonstrative and comparative reference (eg. 'this', 'more'), combined with a drop in explicit endophoric noun phrases with post-head dependents (eg. 'that we did', 'in class'). As the course progresses, special terms and names will level off and there will be an increase in third person personals (eg. 'she'), indefinite pronouns (eg. 'anybody'), exophoric substitution (eg. 'the one') and ellipsis, and superordinates (eg. 'book'), and general nouns and verbs (eg. 'thingy','do'"). This overall trend will be affected by events such as portfolio dates and project deadlines, which lead to conversations laden with technical terms and proper names.

I predict that certain informal aspects of the language will increase as members become more familiar with each other: sociocentric fillers such as 'you know' and 'I mean', slang and expletives, and sentences with no initial subject or auxiliary and unfinished sentences.

I take into account two secondary but essential factors: cohesion and function. A consideration of cohesion should reveal that as the language becomes more exophoric, general and bald, as explicitness is no longer necessary, the risk of breakdowns and requests for clarification increase, especially in course-related topics, because speakers wrongly assume that all the relevant common knowledge is in their hearer's mind.

The analysis of the function of utterances containing cues should show that the use of implicit contextualisation cues is a generally expected unmarked means of claiming in-group membership (Levinson 1978; Tannen 1989). I hypothesise that,

in both transactional and interactional social exchanges, there will be an increase in the students' manipulation of shared knowledge as they flout quality maxims. I predict, too, an increase in speech acts that are demonstrations of 'in-the-same-boatness', such as expressions of dissatisfaction with self or situation, or positive evaluation of the interlocutor.

2.2 Method of data collection

I openly made 15 tape-recordings (a total of 4 hours 40 minutes) of MSc student conversations in the common room of the Applied Linguistics department from 4 October 1991 until 12 May 1992. I recorded once a week over three periods of time: the first half of the first, second and third term. The conversations (29 in total) were spontaneous and unguided, and I kept at a distance at the moment of recording so as not to be included. Six native speakers of English who had options in common and tended to sit together in the common room consistently were, four weeks into the course, selected for analysis, on the assumption that they were representative of all native speakers of English on the course. Once I had transcribed the recordings (26,000 words), I disregarded dialogues or long sections of dialogues which did not contain at least two of the six chosen speakers.

2.3 Method of data analysis

I analysed all discourse units of the text (3,500) in terms of knowledge area and function, in six fields. I also tagged the text (Field seven) lexically and grammatically.

Figure One: Fields and tags of the coding system

<u>Functional Fields</u>:
1) Dialogue and discourse unit numbers
2) Knowledge area
3) Macro-function
4) Move and topic
5) Speech act
6) Speakers

<u>Text Field and linguistic tags</u>
7) Lexical – verbs
nouns
Grammatical articles –
pronouns and adjectives
substitution and ellipsis
dependents

The six functional fields were dialogue number and discourse unit number, knowledge area, move and topic, speech act, and speakers. Of the text tags, the lexical ones were mainly verbs and nouns, tagged with numbers, and the grammatical ones were articles, pronouns and adjectives, substitution and ellipsis, and dependents tagged with letters/symbols.

Let us now examine the individual categorisation of these functional labels and text tags. All examples quoted in this paper are taken from my database. The numbers

beside examples quoted here are from Field one: dialogue and discourse unit. The first two digits indicate the dialogue number; the other three indicate discourse unit number. The letters are from Field six: the speakers. AM, BM, CM, DM were the men selected for study in this research, and AF, BF the women.

3. The Individual Categorisation

3.1 The discourse unit labels

3.1.1 Field one: dialogue and discourse unit.

My criteria for deciding what constituted a discourse unit were as follows: that a unit can end where a grammatical sentence ends, regardless of whether this is the end of a turn or move; that a unit always ends where there is falling intonation combined with a pause longer than 0.5 seconds, even though it happens in the middle of a sentence; and that each discourse unit constitutes a speech act. Any utterance occurring simultaneously, unless it is unintelligible, is considered a unit.

3.1.2 Field two: knowledge area

I divided assumed common knowledge into four areas, in order to distinguish non-course-related topics (areas 1, 2 and 3) from course-related topics (area 4). I added the symbol P if I felt there was shared, privileged or interpersonal knowledge in any of the four knowledge areas.

Figure Two: Knowledge areas

1	The world, Edinburgh
2	Linguistics, language teaching and learning
3	Edinburgh University, Department of Applied Linguistics, Institute for Applied Language Studies,
4	The 91-92 MSc course in Applied Linguistics of Edinburgh University
(P	Privileged or shared knowledge in any of the above four areas)

When I was assigning discourse units to the four knowledge areas, I began to feel that I only needed two categories: course-related and non-course-related. However, since some topics were more course-related than others and some were course-related topics but also related to the 'real world' outside, I maintained the original four areas. Because it was hard to say where one knowledge area ended and another began, at the discourse unit level, a knowledge area was established for the duration of a whole topic, wherever possible.

3.1.3 Field three: macro-function

Discourse units were next labelled T for the transactional function or S for the social, the interactional. I did not predict an increase or decrease in either over time; I wanted to examine the grammatical and lexical tags within the two

functions. Again, the overall function was established for a whole topic, rather than on an individual discourse unit level.

Figure Three: Macro-functions

T	Transactional, instrumental
S	Interactional, social
(F	Flouting of quality and clarity maxims in either of the above two)

T was the purely transactional functional category, the instrumental such as:

e.g. -01054 BM 'Could you get me a tuna and sweet corn one please?'
01055 AM 'Me as well.'

or learning specifically how to do something, transmitting information needed for an immediate task such as:

e.g. -06011 CM 'The best thing to do is use a che- master card or something.'
06012 NF 'I don't have a master card.'

All other units were tagged with S for the social expressive exchange, the interactional category, and the phatic. In many cases, the speakers were testing the normality of their situation or feelings:

e.g. -15020 CM 'But I suppose you're moderately efficient. (1)'
15021 CM 'But I'm not near an outline.'
15022 DM 'Oh God no I'm on my reading.'

In others, they were reassuring themselves that they had the same information and checking each other's attitude to it:

e.g. -08063 CM 'Five's a bad mark right?'
08064 AM 'I think we get five we're fine.'
08065 CM 'It's the other mark I want.'

There is a cline from the instrumental to the purely phatic. Even the purely phatic could have a tinge of information exchange. Sometimes an interaction that, on the surface, was an information exchange could be fundamentally a social exchange.

In the same field, I added a label F to indicate that the speaker was playing with truth and/or the language, flouting cooperative maxims by exaggeration, irony, banter, playing with clichés in mock seriousness, or mixing registers to amuse etc. I wanted to test whether humorous exchanges were a marker of in-groupness that increased over time:

e.g. -04080 BM 'They're sort of we give change!'
04081 DM 'If you can get on the bus you get change then.'
04082 BF 'Ah right // that's good.'
04083 DM '// Yeah.'
04084 AF 'If you don't mind breaking your leg as you try and get on.'

BEST COPY AVAILABLE

3.1.4 Field four: topics and moves

I was especially interested in examining topic shifts to see how often implicit contextualisation cues occurred in the first discourse unit of a new topic. I studied patterns of moves, to see whether this revealed something about group dynamics.

Figure Four: Topics and moves

T	Initiate
R	Respond
1	1st. discourse unit of within a turn, on a new topic
2	Subsequent discourse units within a turn, on an established topic
3	1st. discourse unit of a turn, on an established topic

'Topic' I defined as part of an interaction that can be given a title in the form of an indirect question that covers all speakers' contributions to it. Thus an example of this discourse topic would be 'Why X is not going to do any reading' or 'What Y did on his mountaineering weekend.' To analyse topics, I tagged the first discourse unit of a turn or within a turn on a new topic: '1', I tagged subsequent discourse units within a turn on same topic, even if still answering the same question: '2', and the first discourse unit of a turn on existing topic: '3'. When the topic drift was so gentle that it was almost imperceptible, I placed topic boundaries as close to the centre of the shift as possible.

I classed the moves T for Initiate and R for Respond. R was either the second half of an adjacency pair or simply a minimal response to the preceding discourse, such as a backchannel or prompt to continue. T was everything else. T could occur in T1, T2 and T3. R tended to occur mainly in R3, as in the first discourse unit of an answer or adjacency pair response. R2 occurred rarely because subsequent discourse units of an answer were not normally specifically required by the question.

e.g.	11105	NM 'You went home?'
	11106	AM 'I went home round about half past seven no.'
	11107	AM 'I didn't go.'

11106 was, strictly speaking, all that was needed to answer NM's question. 11107 was T2 rather than R2. Finally, R1 was the change of focus to the metalinguistic, the request for clarification, or the answer to a question from a previous topic, ignoring an interspersed topic.

3.1.5 Field five: speech act

The task of classing discourse units in speech acts was particularly daunting because speech acts are so elusive and one unit often fits into more than one speech act category. When a unit did not fit neatly into an act category, I created a new category, or expanded an existing category, and noted this realisation of the act in order to guarantee consistency.

I initially detected 40 speech acts, but having failed to make them water tight and mutually exclusive, I opted for grouping them into nine macro speech acts with no subdivisions. Discourse units were analysed in terms of attitude (neutral, positive and negative) and then in terms of person or object discussed (self, interlocutor or communication, and third party or situation). This solution obviously did not produce a delicate tool but I felt that it would be more manageable for the quantitative analysis stage.

Figure Five: Speech acts

attitude towards:	Neutral	Positive	Negative
Self	11 I do/am	12 I do/am good	13 I do/am bad
Interlocutor/ communication	21 You do/are Here we are	22 You're good/ I'm with you	23 You're bad/ I'm not with you
Third party/ situation	31 It/she does/is	32 It/she does/is good	33 It/she does/is bad

This paper permits no more than a general view of each macro speech act. As far as Attitude towards Self was concerned, within act 11, I included inform, explain, answer and express own intentions and desires; within act 12, express satisfaction with self, and reassure and console self; within act 13, express dissatisfaction with self, minimise praise of self, excuse self and apologise.

Similarly, for Attitude towards Interlocutor or Communication, there was in act 21, greet, fill phatically with afterthoughts, backchannelling, laughter etc., request, in act 22 evaluate interlocutor positively, console, encourage, sympathise, agree, predict drift, advise and suggest, offer/promise goods or action, and in act 23, challenge factual content/truth value, decline, and deny.

Finally, in Attitude towards Third Party or Situation, I included in act 31, inform, explain, ask/answer; in act 32, express satisfaction with third party or situation; and in act 33, express dissatisfaction with them, express fear and apprehension. And thus I hoped to cover everything that MSc students do with language in the common room.

3.2 Field seven: the tagged text

3.2.1 Lexical tags

Although the main emphasis of my study is the noun phrase, I included an analysis of all verbs, because I wanted to show that over time there would be an increase in general verbs 'do' and 'make', in terms of a percentage out of all verbs.

e.g. -10069 CM "I've <u>done</u> all the people"

I had one category for the course-related general verb and one for the non course-related.

The other category of verb that I tagged separately from the 'all verbs' category was the filler: both sociocentric sequences (Stubbs 1983) such as 'I mean', 'You see', and omissible hedging verbs with or without sentential objects (Brown and Miller 1980) such as 'I suppose.'

e.g. -17020 AF 'It is a lot you know three thousand words'.

Figure Six: Lexical tags - verbs

10	All Verbs (excluding be, filler, general verbs)	eg: 'bought'
11	Be	eg: 'is'
12	Filler	eg: 'I mean'
13	Non-course general verb	eg: 'do'
14	Course general verb	eg: 'do'

In my categorisation of nouns, I tagged common nouns, proper nouns and general nouns, indicating whether the referent was course-related or non-course-related.

I tagged expletives and slang, being interested in these as markers of intimacy that might increase in number with repeated interaction. I included certain adjectives:

e.g. -08020 AM 'Bloody tosh, isn't it?'
08021 CM 'Well it's- it's a bit abstract. (11)'

within the group of expletives, and phrasal expressions:

e.g. -10029 DM 'So what on earth is it going to be about?'

within the slang group.

Figure Seven: Lexical tags - nouns

''	Expletives	eg: 'hell'
''	Slang	eg: 'scivers'
20-22	Non-course nouns	
20	Common non-course nouns	eg: 'budgie'
21	Proper non-course names	eg: 'Japan'
22	General non-course nouns	eg: 'thing'
30-50	Course nouns - special	
30-32	Common special course nouns	
30	technical	eg: 'X-bar'
31	unique	eg: 'portfolio'
32	superordinate	eg: '[syntax] book'
40-41	Proper special course names	
40	actual use	eg: 'Chomsky'
41	metonymical use	eg: 'Chomsky [study]'
50	General special course nouns	eg: 'thing'
60	Course nouns - course-by-context	eg: 'work [for the exam]'

Categorising nouns with non-course-related referents was quite simple. Words such as 'budgie' and 'bus-stop' were classed as common nouns, 'Bush' and 'Manchester' as proper nouns, and 'thing' and 'people' as general nouns.

Classing nouns with course-related referents was more complex. These were divided into two main groups: special course nouns and course-by-context nouns. The special course nouns, intrinsically course-related, were classed under the headings common, proper and general nouns.

Under the first special course nouns heading, that of common nouns, I made three sub-divisions - technical, unique and superordinate. I tagged nouns technical if they were intrinsically specialised terms independent of the context of the course, linguistics and language teaching theory terminology such as 'X-bar', 'diglossia' and 'lexical syllabus'. I tagged nouns unique common special course nouns if they were single course components such as 'portfolio', 'core project', and 'the examiner', used in department organisation, the course handbook etc. Superordinate common special course nouns constituted a class of many possible course components. Superordinates were count nouns such as '[syntax] book', '[SLA] class' and '[tutorial] task' whose precise meaning was not clear since they were the second noun of a two-word phrasal expression (Huddleston 1988:103) whose first word (usually a nominal pre-head modifier) was omitted. In this

e.g. -20020 CM 'SLA class next week and the paper's due in next Friday'

the paper was the core project, the first project paper of the course.

Under the proper special course noun heading there were two sub-divisions: one for proper nouns with their actual intended use, such as 'Chomsky', 'ELT' and 'Structuralism', and the other for their metonymical use, in which the noun referred elliptically to something other than the course/linguist named. As with the superordinate common special nouns, these nouns could be seen as one word of a two-word phrasal expression, except that in the case of the metonymical use, it is the second word that is omitted. In

e.g. -08031 AM 'Though though I haven't I haven't done any Chomsky'

'Chomsky' meant Chomsky study or revision for the forthcoming examination.

The general special course noun category was structurally the same as the general non-course noun. To take an example:

e.g. -10069 CM 'No I've done all the people'

Here, 'people' meant theorists or linguists.

The second main group of course nouns I called course-by-context nouns. These were not intrinsically course-related but became course-related by their use in course-related topics. The referents of 'discussion', 'work' and 'this week' were course-related, for instance, when the 'discussion' was one that the students had had in a tutorial, 'work' meant work for the examination, or 'this week' meant this

BEST COPY AVAILABLE

week in classes. Here what seemed to be missing was a post-head dependent such as a relative clause, a prepositional phrase or a content clause.

3.2.2 Grammatical tags

Grammatical tags were put on articles, pronouns and adjectives, substitution and ellipsis, and dependents.

Figure Eight: Grammatical tags - articles, pronouns, adjectives, substitution and ellipsis

Articles		
T/A	indefinite article / definite article	
Pronouns and adjectives		
P	Indefinite pronoun	eg 'anybody'
E	Existential pronoun	eg 'she'
S	Possessive adjective	eg 'their'
D	Demonstrative pronoun / adjective	eg 'this'
C	Comparative pronoun / adjective	eg 'more'

I tagged all articles, predicting that there would be an increase in the partially explicit exophoric use definite noun phrase, especially with course-related nouns (including general course-nouns), as in this example:

e.g. - 16032 AF 'I think- I find a real loss actually of not having read the (0.5) Fay Cutler article which seems to be underpinning this ='

I tagged the indefinite pronouns 'everything', 'nobody' etc., feeling that they have an exophoric quality. They constituted a 'vague' feature of in-group language. In the following example:

e.g. - 25020 NF 'Is there anybody else applying?'
the full extent of 'anybody else' could only be understood by group members.

The tagging of pronouns, adjectives, substitution and ellipsis proved to be a complex task. There were different degrees of endophora and exophora. I considered pronouns and adjectives endophoric if the referent was in the text, whether it was an item or a proposition, even though the reference would have been unclear or ambiguous to our hypothetical outsider because of the existence of a number of possible textual referents. I decided to tag as semi-implicit endophoric, the adjectives that modified general nouns and the pronouns and substitutes that had endophoric cohesion to general nouns, exophoric pronouns or substitutes, or indefinite pronouns. On the borderline between endophora and exophora seemed to be associative anaphora in which the referent could be understood from the schemata established by other referents in the text.

The rest of the pronouns, adjectives, substitutes and ellipses were exophoric or non-anaphoric. Here, the degrees of implicitness were more noticeable. Again I differentiated between adjectives that modified common nouns explicitly, and those that could be called semi-implicit exophoric because they modified general nouns

or exophoric substitutes. Most demonstratives, for example, are exophoric, but demonstrative adjectives seemed more implicitly exophoric when they modified the impenetrable, intertextual general nouns or exophoric substitutes.

e.g. 15045 DM 'So I typed that thing up again after you'd gone.'

The pronouns, substitutes and ellipses whose referents were irrecoverably right outside the text were classed as totally implicit exophoric.

Tagging pronoun and adjective categories revealed interesting particularities of each. Demonstratives seemed more obviously intertextual than other forms of reference. In the following example, the 'pizza' outing had evidently been discussed on a former occasion, yet it was exophoric as far as the text was concerned.

e.g. 11089 AM 'Did you go to this pizza on Friday?'

Some of the characteristics of demonstratives were typical of spoken language. I tagged the demonstrative pronoun in the expression 'like that' in an unfinished sentence not as a demonstrative but as a substitute expression, or 'vague expression' (Channell 1985). Familiarity with the presuppositional pool is needed to complete the sentence.

e.g. 04089 BF 'And King's Building's got a bar and stuff like that.'

Narrative sections featured the informal demonstrative adjective 'this', which brings the story closer to the hearers.

e.g. 02084 MM 'So so he keeps he's got this really incredible part.'

Comparative reference has three points of reference: the referent mentioned, the referent that it is compared to, and the quality that unites or divides them. I was tempted to tag some of the cases in my data exophoric, because of the exophoric comparing quality.

e.g. 15076 DM 'The actual thing's exactly the same.'

I opted not to call these cases exophoric, however, since the 'exophoric-ness' of a quality was too flimsy to pin down.

Two final points need to be made in connection with tagging pronouns and adjectives in spontaneous recordings of groups of three or more speakers. The first is that since few recordings contained beginnings of conversations, most conversations were captured mid-topic. Thus there were reference items that looked as if they might have been endophoric, with the referent probably in the preceding, unrecorded, stretch of conversation. Faced with this uncertainty, I tagged such cases using guesswork based on the text.

The second point is that some speakers missed the textual referents because they joined conversations already under way. Thus, although I might tag one item of reference as endophoric, it could seem exophoric to some hearers. Witness

e.g. 11064 BF 'And they were doing a profile of him.'
11065 BF 'And I was sort of =.'
11066 NM 'Who was it?'
11067 BF 'He was a guy who wrote this dirty book.'
11068 BF 'Melvyn Bragg?'

Here, BF had to supply the missing restrictive post-head dependent and proper name, to repair the breakdown.

Tagging substitution and ellipsis, I found that exophoric substitution was predictably rare, apart from the vague substitute expression at the end of the unfinished sentence, and that ellipsis was more interesting to tag as there were more instances of implicit use. Here is an example of semi-implicit endophoric: the 'them' in line 25014 is exophoric.

e.g. 25014 CM 'So what are em how many of them are there do you know?'
25015 BF 'I don't know. (0.5)'
25016 BF 'Don't know.'
25017 BF 'Em.'
25018 BF 'I think (1) three or two. I'm not sure.'

I tagged ellipsis of initial subject or auxiliary, interested in it as an example of speakers' inexplicitness about the "given" part of the information structure. This informal style could be a marker of intimacy.

e.g. 24040 DM 'Something wrong isn't there somewhere?'

The ultimate ellipsis is the unfinished sentence: the speaker assumes that the hearer can supply the missing information.

e.g. 26131 BM '// Have you got this on the the drive at the,,?'
26132 CM 'No // no you can't do that.'

I call an unfinished sentence one that is deliberately left 'in the air', and not one that is unfinished because another speaker interrupts.

The last grammatical elements, dependents, were fairly straightforward to tag. I predicted that as assumed common knowledge increased, there would be less need to identify the nouns with a description, whether in the form of a pre-head dependent (determiner or modifier) or a post-head dependent (complement, modifier and peripheral). Post-head dependents will abound in the first half of the course:

eg: 16088 AF 'I mean it all to do depends on the point of view of the research which we're not really // adequate to judge.'

whereas, by the end, the hearer will probably have to supply them:

eg: 27128 DM 'If you're answering questions anyway so (3) shou'dn't be a problem. ((2))'

Figure Nine: Grammatical tags - dependents

Pre-head	determiner, modifier
Post-head	complement, modifier, peripheral

4. Conclusion

4.1 Summary and comments

This paper has examined a coding system for analysing the recorded casual conversations collected in the 1991-92 Applied Linguistics MSc common room. It has explained the system of six functional fields to analyse the speech acts and knowledge areas of each discourse unit, and one actual text field with lexical and grammatical tags. The problems of each have been detailed and the solutions described.

Although I first constructed the coding system using the information that I had gained from an impressionistic global survey of the data, once I was actually using the code to analyse the dialogues, I had to adjust it constantly in response to the dictates of the data. On finishing the coding, I had to adapt and re-shape the system again to make all the categories fit together into a coherent model. Finally, I have confirmed the workability of the code, using inter-coder reliability tests.

4.2 The next stage in the research

Using the Excel database, the next stage is to count the implicit contextualisation cues and other linguistic items, as well as the discourse units in the function fields of knowledge area, macro-function, topic shifts and moves, and speech acts, in each dialogue, and to find the average percentage of these for each of the three recording periods of the course, and thus observe whether there are any obvious changes over time: whether knowledge area four topics do increase, whether the number of cues does increase, which cues increase at which point, whether language becomes more informal, and so on. The second stage is to carry out a qualitative analysis of chunks of dialogue or even whole dialogues to find characteristics of the ways that cues relate to each other within a text.

The third stage is to put all the information together, to observe linguistic items in conjunction with the various functional fields and discover whether their density varies according to the knowledge area, the macro-function and the speech act. This would test the hypotheses that course-related topics have more implicit, impenetrable language, and that the use of cues is related to the claiming of in-group membership. Impenetrability itself will be tested through questionnaires given to speakers from each of the four knowledge areas.

The subsequent statistical analysis should help me to select the most significant implicit contextualisation cues to include in the final model which should suit any body of data taken from conversations between in-group members of an academic discourse community. I hope eventually to make the model generalisable to English-speaking academic discourse communities in Britain, and possibly non-

BEST COPY AVAILABLE

academic discourse communities to explain why their conversations can sometimes be impenetrable to outsiders.

References

Bernstein B. 1971. Class, Codes and Control. Vol 1. London: Routledge and Kegan Paul.

Brown E. and J. Miller. 1980. Syntax: A Linguistic Introduction to Sentence Structure. London: Hutcheson University Library.

Brown P. and S. Levinson. 1978. Politeness. Cambridge: CUP.

Channell J. 1985. 'Vagueness as a conversational strategy' in Nottingham Linguistic Circular. 14: 3-24.

Duck S. 1991. Friends For Life. Hertfordshire: Harvester Wheatsheaf.

Gumperz J. 1982. Discourse Strategies. Cambridge: CUP.

Halliday M. and R. Hasan. 1976. Cohesion In English. London: Longman.

Huddleston R. 1988. English Grammar: An Outline. Cambridge: CUP.

Kreckel M. 1981. Communicative Acts and Shared Knowledge in Natural Discourse. London: Academic Press.

Levinson S. 1983. Pragmatics. Cambridge: CUP.

Levy D. 1979. 'Communicative goals and strategies: between discourse and syntax' in T. Givon (ed.) 1979. Syntax and Semantics 12:183-210.

McCarthy M. 1991. Discourse Analysis for Language Teachers. Cambridge: CUP.

Stubbs M. 1983. Discourse Analysis. Oxford: Basil Blackwell.

Swales J. 1990. Genre Analysis. Cambridge: CUP.

Tannen D. 1989. Talking Voices. Cambridge: CUP.